For Alzheimer disease GWAS, pulling needles from the haystack is just the first step

Russell H. Swerdlow,
MD

Elizabeth H. Corder,
PhD

Correspondence & reprint
requests to Dr. Swerdlow:
rswerdlow@kumc.edu

Neurology® 2012;79:204–205

Late-onset Alzheimer disease (LOAD) does not demonstrate classic Mendelian inheritance patterns and is considered a sporadic disorder. Inheritance, though, influences LOAD risk. Single nucleotide polymorphisms (SNPs) in the *APOE* gene associate with AD, and LOAD *APOE4* carriers overall develop symptoms at an earlier age than noncarriers.[1,2] Various studies suggest that mitochondrial DNA, which does not follow Mendelian rules, also plays a role.[3] A multitude of association studies have tested SNPs in individual candidate genes with variable results.[4] Genome-wide association studies (GWAS), which simultaneously interrogate many SNPs, suggest a number of genes influence LOAD risk.[5–9]

In a GWAS, DNA from subjects with and without a disease is hybridized to chips that genotype 300,000 to 1 million of the 14 million currently validated human SNPs. Because SNPs in a particular region frequently predict other nearby SNPs, GWAS chips can efficiently determine much of the common DNA variation found across an individual's 3 billion base pair genome and 20,000–25,000 genes. When the frequency of a particular SNP base differs between disease-affected and control subjects, it infers a gene within that SNP's region influences disease risk.

Comparing the frequency of SNP variants in AD and unaffected control subjects, though, provides only partial insight into LOAD biological processes. In this issue of *Neurology*®, Allen et al.[10] help to address this limitation. The authors postulated at least some recently identified LOAD-associated SNPs could affect LOAD risk through effects on gene expression (eSNPs). To test this they determined, in autopsy brains, mRNA expression levels of genes located close to GWAS-implicated loci. They asked whether the expression of these genes related to AD status, and if genetic association could be used to identify nearby eSNPs.

Functional variation was found in 2 instances. SNPs in the CLU (rs11136000) and MS4A4A (rs2304933/rs2304935) genes influenced their expression levels in temporal cortex. The LOAD-protective CLU and the risky MS4A4A alleles both occurred in conjunction with elevated levels of expression. An additional CLU SNP was also associated with both CLU expression and LOAD, provisionally qualifying it as a newly identified LOAD eSNP. Although GWAS-implicated ATP-binding cassette subfamily A member 7 (ABCA7) SNPs did not appear to function as eSNPs, ABCA7 expression levels differed between the LOAD and control groups and a subsequent search for ABCA7 eSNPs revealed several potential LOAD-associated candidates.

If correct, the findings of Allen et al. could advance the LOAD research field. Determining why and how CLU, MS4A4A, and ABCA7 expression influences LOAD risk may provide mechanistic insight into LOAD biology. Drugs that modify their protein products or the pathways these products contribute to could theoretically modify LOAD risk or progression.

GWAS and gene expression arrays generate enormous numbers of data points, which increases the chance of reporting false-positive associations. In this case it is reassuring that gene expression changes occurred in conjunction with risk loci. Replication, though, is still important to ensure the observed relationships were not spurious. Since RNA was obtained from autopsy brains it is important to consider whether disease stage, agonal state, or postmortem factors could have influenced results. Finally, it should be noted that differences in gene expression do not rule out the possibility that other processes can also occur, such as protein primary structure changes, either in the gene of interest or in other nearby genes.

While GWAS have effectively pulled genetic needles from the LOAD haystack, collecting these needles is only a first step. In this respect the study of Allen et al. represents a welcome fundamental step forward. It presents a means through which statistical

associations can be confirmed at a functional level and used to understand the biological basis of LOAD.

DISCLOSURE

REFERENCES

1. Corder EH, Saunders AM, Strittmatter WJ, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science 1993;261: 921–923.
2. Blacker D, Tanzi RE. The genetics of Alzheimer disease: current status and future prospects. Arch Neurol 1998;55: 294–296.
3. Swerdlow RH. Mitochondria and cell bioenergetics: increasingly recognized components and a possible etiologic cause of Alzheimer's disease. Antioxid Redox Signal Epub 2011.
4. Bertram L. Alzheimer's disease genetics current status and future perspectives. Int Rev Neurobiol 2009;84:167–184.
5. Harold D, Abraham R, Hollingworth P, et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. Nat Genet 2009;41:1088–1093.
6. Seshadri S, Fitzpatrick AL, Ikram MA, et al. Genome-wide analysis of genetic loci associated with Alzheimer disease. JAMA 2010;303:1832–1840.
7. Lambert JC, Heath S, Even G, et al. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. Nat Genet 2009;41:1094–1099.
8. Hollingworth P, Harold D, Sims R, et al. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. Nat Genet 2011;43:429–435.
9. Naj AC, Jun G, Beecham GW, et al. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. Nat Genet 2011; 43:436–441.
10. Allen M, Zou F, Chai HS, et al. Novel late-onset Alzheimer disease loci variants associate with brain gene expression. Neurology 2012;79:221–228.